1-1-2015

Topological Complexity in Protein Structures

Erica Flapan
Pomona College

Gabriella Heller '14
Pomona College

Recommended Citation

E. Flapan, G. Heller, Topological Complexity in Protein Structures, Molecular Based Mathematical Biology, Vol 3, Issue 1 (2015) 23–42.

Research Article **Open Access**

Erica Flapan* and Gabriella Heller

Topological Complexity in Protein Structures

Abstract: For DNA molecules, topological complexity occurs exclusively as the result of knotting or linking of the polynucleotide backbone. By contrast, while a few knots and links have been found within the polypeptide backbones of some protein structures, non-planarity can also result from the connectivity between a polypeptide chain and inter- and intra-chain linking via cofactors and disulfide bonds. In this article, we survey the known types of knots, links, and non-planar graphs in protein structures with and without including such bonds and cofactors. Then we present new examples of protein structures containing Möbius ladders and other non-planar graphs as a result of these cofactors. Finally, we propose hypothetical structures illustrating specific disulfide connectivities that would result in the key ring link, the Whitehead link and the 5_1 knot, the latter two of which have thus far not been identified within protein structures.

Keywords: non-planarity; proteins; topology; knots; links; spatial graphs; Möbius ladders

MSC: 57M25, 92C40

DOI 10.1515/mlbmb-2015-0002
Received February 8, 2014; accepted January 16, 2015

1 Introduction

Until two decades ago, the underlying form of all native protein structures was assumed to be topologically linear. This was in contrast to DNA molecules whose polynucleotide backbones had been known for a decade to exist in both linear and closed circular knotted and linked forms [4, 6, 40]. In 1994 and 1995, Liang and Mislow [9–11] showed that if cofactors and disulfide bonds are taken into account when evaluating the topological form of protein structures, then these structures may indeed contain knots, links, and even non-planar graphs. More recently, even when considering only the polypeptide backbones of proteins, some knots and links have been identified [3, 13, 15, 23, 32, 33, 36].

Why some proteins contain these non-planar features has been a question since their discovery. The fact that features such as knots and non-planar spatial graphs have been highly preserved throughout evolution suggests that they may play an important role in enzymatic activity. For example, while ubiquitin hydrolase has only a 33% sequence homology between its human and yeast forms, it contains the same five-fold knot within the respective structures [28]. Interestingly, Virnau et al. [36] noted a pair of homologues for which the topology of the backbone was not preserved: N-acteylornithine transcarboamylase (AOTCase) and ornithine transcarboamylase(OTCase). AOTCase is an enzyme used for arginine biosynthesis in major pathogens. Other bacteria, animals, and humans utilize OTCase to alternatively process L-ornithine. A deep knot in an active site of AOTCase, which is absent in OTCase, enables the AOTCase to bind acetylornithine, while OTCase will bind only L-ornithine. It is believed that the rigidity of the knot restricts ligand access to any buried residues, enabling the interaction of acetylornithine with those that are strictly surface-exposed. The flexibility from

***Corresponding Author: Erica Flapan:** Department of Mathematics, Pomona College, 640 North College Avenue, Claremont, CA 91711 USA, E-mail: erica.flapan@pomona.edu
Gabriella Heller: Department of Mathematics, Pomona College, 640 North College Avenue, Claremont, CA 91711 USA, E-mail: gth22010@mymail.pomona.edu

the absent knot in OTCase permits greater access to its respective ligand, and demonstrates how the presence of a knot can greatly affect the active site and function of otherwise homologous proteins [36].

In this paper we study the topological complexity of the underlying structures of proteins. In the field of structural biology, the term "topology" can be used in different ways. However, in this paper, we use the word “topology” in the mathematical sense, which considers two objects to be topologically equivalent if one can be deformed to the other in space. In particular, we define an object to be *topologically complex* if it is non-planar in the sense that even assuming complete flexibility and elasticity, it cannot be deformed into a plane. For example, a knotted circle is non-planar and hence is indeed a type of topological complexity. Understanding how such non-planarity arises in proteins may offer valuable insight into protein folding mechanisms and degradation pathways.

In order to evaluate their topological complexity, we model protein structures as completely flexible objects. While a protein molecule is in fact only partially flexible about its peptide bonds, if we can show that a completely flexible model cannot be deformed into a plane, then the protein, with less flexibility than the model, also cannot attain a planar conformation. Thus, for the purpose of identifying non-planarity within protein structures, we can assume complete flexibility, even though this is not physically accurate. Figure 1 illustrates how, for a given a protein with an organized tertiary structure, we associate a flexible model that we refer to as its *underlying topological structure*. Because the backbone of the protein in the figure contains neither disulfide bonds nor other cofactors connecting non-adjacent regions, its underlying topological structure is an open linear path.

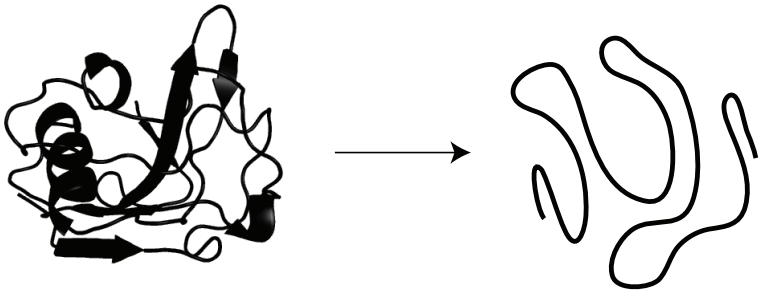

Figure 1: We represent the protein ribbon diagram on the left by a flexible underlying topological structure, shown on the right.

In Section 2, we present a survey of non-planar topological structures that have been previously identified within proteins. Then in Section 3, we present new examples of proteins containing the non-planar graphs $K_{3,3}$ and K_5 as well as some containing Möbius ladders as a result of covalently-bound cofactors. Finally, in Section 4, we propose hypothetical protein structures whose covalent bonding would result in a 3-component key ring link, the Whitehead link, and the 5_1 knot, the latter two of which have not been previously identified in the underlying structures of proteins.

2 Survey of Known Non-planar Protein Forms

Knots

Mathematicians define a *knot* as a circular path in three dimensional space. See, for example, the image on the left side of Figure 2. The requirement that the path be circular is so that the knot is topologically trapped within the figure. By contrast, a knot which is contained in an open linear segment (as shown on the right in Figure 2) can be removed by a deformation of the figure.

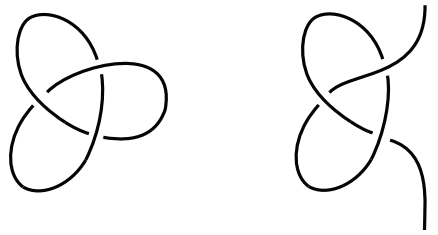

Figure 2: The knot on the left is trapped in a circular path, whereas the knot on the right can be unknotted by a deformation.

While the polypeptide backbones of protein structures are normally open segments rather than circular paths, if a knot is deeply embedded in a protein structure, then from a biochemical viewpoint, it is reasonable to assume it is trapped within the structure. This assumption generally holds as a result of the knot's location in a valley of the potential energy landscape. In particular, the energy required for the termini, usually near the protein surface, to unthread the knot is so prohibitively large that this event generally does not occur. However, it is a difficult problem to create a model of such a knotted open segment in which the knot is topologically trapped within the model and can be uniquely identified.

Many approaches have been developed to create such a topological model [1, 7, 14, 22, 30, 35]. One approach is to extend the termini indefinitely in opposite directions [8]. We illustrate this by adding arrows at the ends of the linear segment as shown in Figure 3. However, if one or both of the termini are near the tangling of the knot, then different knots can be trapped in the structure depending on how the ends are extended [17, 19, 25]. For example, in the right image of Figure 3, if we extend the bottom endpoint under the arc we will get a trefoil knot, but if we extend it over the arc we will get the unknot.

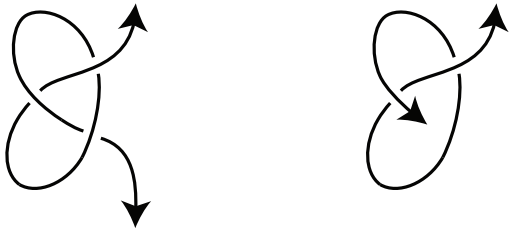

Figure 3: For the image on the left we can extend the endpoints of the segment indefinitely so that the knot is trapped and is uniquely determined. But for the image on the right, we may obtain either a knot or an unknot depending on how we extend the bottom endpoint.

Another approach to modeling a knotted linear protein structure is to place it in the center of a very large ball. The termini are then extend to the spherical boundary of the ball and joined by an arc on that boundary to create a closed loop [14, 30]. However, the same problem can occur that we saw on the right side of Figure 3 when we extended the ends indefinitely. In particular, different knots may be obtained depending on how the arcs are extended to the boundary of the ball. To resolve this problem, Millett and others [12, 16, 18, 20, 21] have proposed a statistical approach in which they assign different probabilities to each of the possible knots they obtain in this way. Then, only the knot with the highest probability is considered to occur in the given protein structure.

Whatever method is taken to modeling and characterizing knots in proteins, amongst the tens of thousands of known structures, only a few hundred have been found to contain knots. Furthermore, every one of the known knotted proteins is deformable to one of the knots 3_1, 4_1, 5_2 and 6_1, illustrated in their circular forms in Figure 4.

Topologists refer to every knot by a number together with a subscript. The first number indicates the minimum number of crossings required in any illustration of the knot. Since this number does not necessarily uniquely determine the knot, the subscript is then used to distinguish knots with the same minimum number

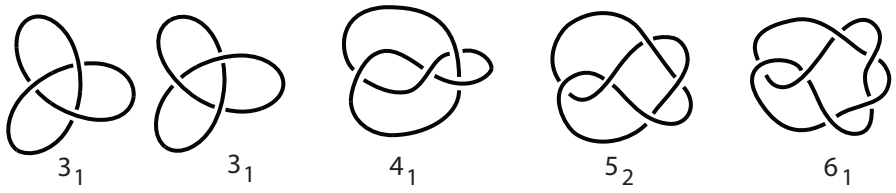

Figure 4: These are the only knot types that have been identified in protein structures to date.

of crossings. While the trefoil knot 3_1 and the figure eight knot 4_1 are the only knots containing three and four crossings respectively, there are two distinct knots 5_1 and 5_2 containing five crossings and three distinct knots containing six crossings. For knots which are topologically distinct from their mirror image, the same pair of numbers is used to refer to both the knot and its mirror image. Of the knots illustrated in Figure 4, the figure eight knot 4_1 is the only one which is topologically equivalent to its mirror image. Figure 4 illustrates both mirror forms of the 3_1 knot, since both have been identified in proteins. The 5_2 and 6_1 knots have each only been found in proteins in the forms illustrated in Figure 4 (see [24]).

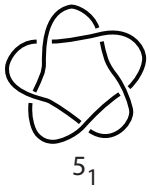

Figure 5: This five crossing knot that has yet to be found in a protein structure.

It is perhaps surprising that the relatively simple knot 5_1 (illustrated in Figure 5) has yet to be found in any protein structure. Taylor has hypothesized that this knot is an unlikely motif in protein structures, because a rate limiting "double threading" event would be required for its formation [34]. Taylor contrasts this with the 5_2 knot, which could be the result of threading one end of the backbone through a twisted hairpin, and then connecting the termini with an arc to create a closed loop as in the model introduced by Millet and others (discussed above). We illustrate such a threading in Figure 6. In Section 4, we discuss how the 5_1 knot may hypothetically occur in a protein structure with and without considering disulfide bonds.

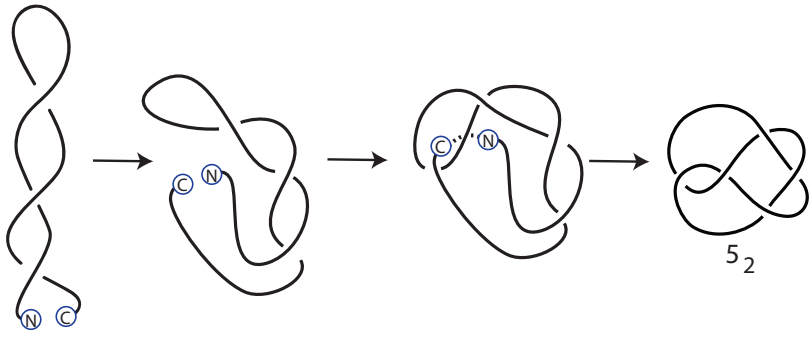

Figure 6: The 5_2 knot could be the result of threading one end of the backbone through a twisted hairpin as illustrated.

The first knotted protein structures to be identified were the metalloproteins ascorbate oxidase and human lactoferrin, which were found by Liang and Mislow to contain the trefoil knot 3_1 [9]. The existence of these knots relied on the covalently bound metal atoms and disulfide bonds of the protein. For example, we

can see in Figure 7 that the trefoil knot found in ascorbate oxidase makes use of the thick grey segments representing disulfide bonds.

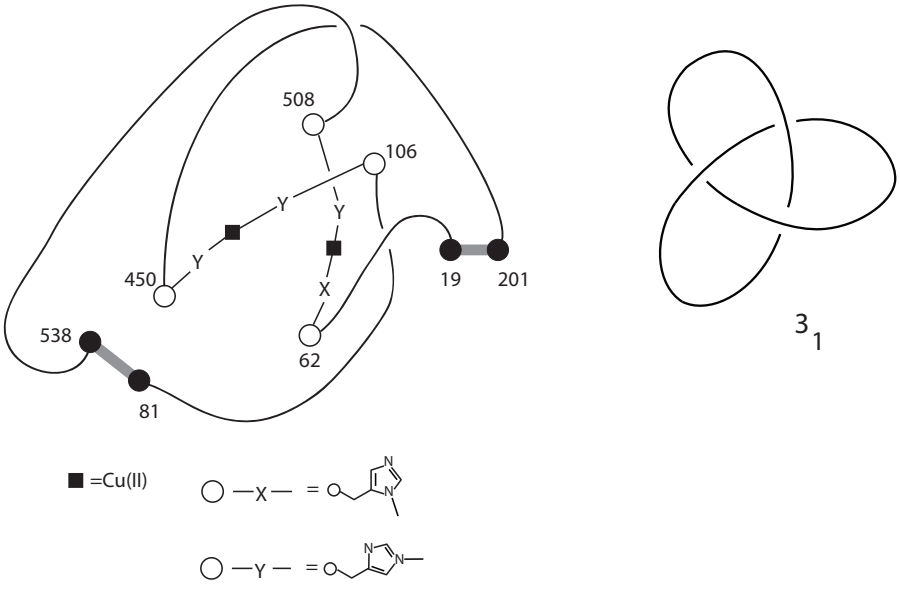

Figure 7: The trefoil knot contained in ascorbate oxidase makes use of the thick grey segments representing disulfide bonds.

More recently, nearly forty proteins have been identified which have knots entirely formed by their polypeptide backbones. The trefoil knot is by far the most common knotted motif reported in protein backbones. Both the left- and right-handed forms of the trefoil have been identified. Right-handed trefoils include carbonic anhydrases and S-adenoylsmethionine synthetase, while ubiquitin is an example of a protein containing a left-handed trefoil [33]. Generally, trefoil knots are highly preserved between species. For example, the carbonic anhydrase trefoil is conserved in isozymes ranging from bacteria and algae to humans, perhaps suggesting its significance in biological function. By contrast, only a few proteins are currently known to contain the figure eight knot 4_1. [36].

In 2006, the first knot with five crossings was identified in the backbone of a protein by Virnau et. al. This 5_2 knot is contained in ubiquitin hydrolase, a protein responsible for rescuing proteins from degradation. It is believed that the complex knotted structure of this protein helps protect it against degradation. This theory is supported by the fact that protein degradation is generally initiated by an ATP-dependent event in which proteins are unfolded by being threaded through a narrow pore of a proteosome, of about 13 Å in diameter. Thus, it is plausible that a knot can sterically hinder such threading, and hence prevent unfolding and degradation [36]. Note that further experiments need to be carried out to verify this. The most complex protein knot that has been found to date is a Stevedore's knot 6_1, which was identified in a α-haloacid dehalogenase structure in 2010 [2].

Table 1 gives a list of the knots, links, and non-planar graphs found in proteins to date and a brief description of how they arise (e.g., from the backbone, from cofactors, etc.). Note that the table provides a summary of the types of topological diversity that have been identified in protein structures, but does not provide a comprehensive list of all proteins that contain non-planar features. More extensive lists of knotted proteins can be found elsewhere [24, 33, 36, 42].

Table 1: Selected Examples of Topological Complexity Previously Found in Protein Structures

	Selected Examples	Citations
Knots		
3_1	**cofactors used:** ascorbate oxidase, human lactoferrin **backbone only:** hypothetical protein, plasmid pTiC58 VirC2, N-succinyl-L-ornithine transcarbamylase, methyltransferase domain of human TAR RNA binding protein, alpha subunit of human S-adenosyl-methionine synthetase, human carbonic anhydrase II, ribosomal 80S-eEF2-sordarin complex	[9, 11, 24, 32, 36, 42]
4_1	**backbone only:** acetohydroxy acid isomeroreductase, photosensory core domain of aeruginosa bacteriophytochrome	[24, 33, 42]
5_2	**backbone only:** ubiquitin hydrolase	[24, 42]
6_1	**backbone only:** α-haloacid dehalogenase	[24]
Links		
Hopf	dimeric citrate synthase from P. aerophilum (two subunits are linked by two intramolecular disulfide bonds) bovine mitrochondrial peroxiredoxin III (two interlinked rings, each consisting of twelve subunits) tuna cytochrome c (the porphyrin makes up one ring and the protein backbone together with central heme iron atom make up another).	[3, 11, 42]
Key Ring	Cytochrome c3 from Desulfovibrio vulgarism Miyazaki	[11]
Chain Mail	bateriophage,HK97 (72-component linked chain mail)	[41, 42]
Non-Planar Graphs		
$K_{3,3}$	Fe4S4 cluster of chromatium high potential iron protein	[10]
K_5	Fe4S4 cluster of chromatium high potential iron protein	[10]

Links

Recall that mathematicians define a *knot* as a circular path in three dimensional space. We say two or more disjoint circular paths in three dimensional space are *linked*, if there is no way to deform the paths so that at least one of the them lies on either side of a plane. For example, the pair of circles on the left in Figure 8 is linked, whereas the pair of circles on the right is unlinked because it can be deformed so that one circle is on either side of a vertical plane. The most common type of protein link has the form of the *Hopf link*, which consists of two circles linked together only once as illustrated on the left of Figure 8.

The Hopf link has been found in protein structures in several distinct ways including the linking of two polypeptide chains via intramolecular disulfide bonds in dimeric citrate synthase from *P. aerophilum*, the linking of multi-subunit macromolecular non-covalent rings in bovine mitrochondrial peroxiredoxin III [3, 42], and linking resulting from the inclusion of porphyrin cofactors in tuna cytochrome c [11]. These three different forms of protein Hopf links are illustrated in Figure 9.

Other topologically distinct links can be found as a result of multiple heme cofactor binding. In particular, by treating each of its heme cofactors as its own ring, cytochrome c_3 from *Desulfovibrio vulgarism* Miyazaki

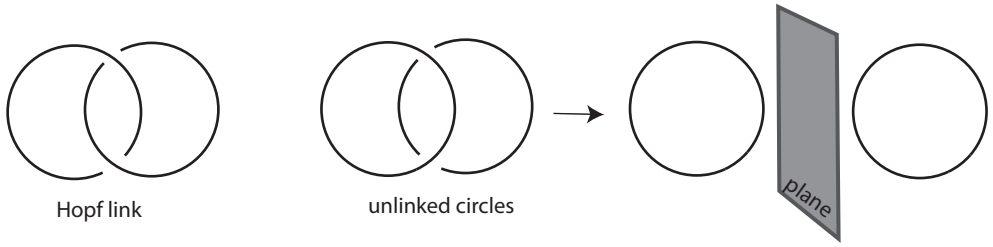

Figure 8: The pair of circles on the left is a Hopf link, but the pair of circles on the right is unlinked.

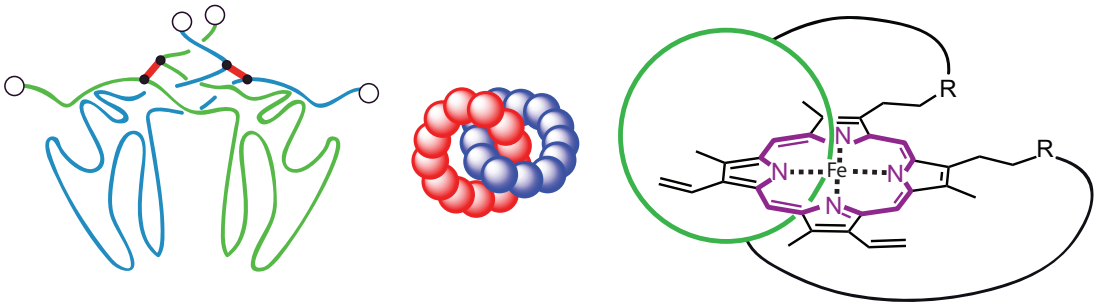

Figure 9: Examples of Hopf links in protein structures. Left: dimeric citrate synthase from *P. aerophilum* in which two subunits are linked as a result of intramolecular disulfide bonds, shown as thick red lines. Center: bovine mitrochondrial peroxiredoxin III in which two interlinked rings, each consisting of twelve subunits, form the Hopf link. Right: tuna cytochrome c in which the porphyrin makes up one ring (purple ring containing nitrogen atoms and double bonds) and the protein backbone together with central heme iron atom make up the other ring (green).

can be seen to contain a five-component link [9]. In particular, by omitting the pink arcs in Figure 10, we obtain the key ring link illustrated on the right.

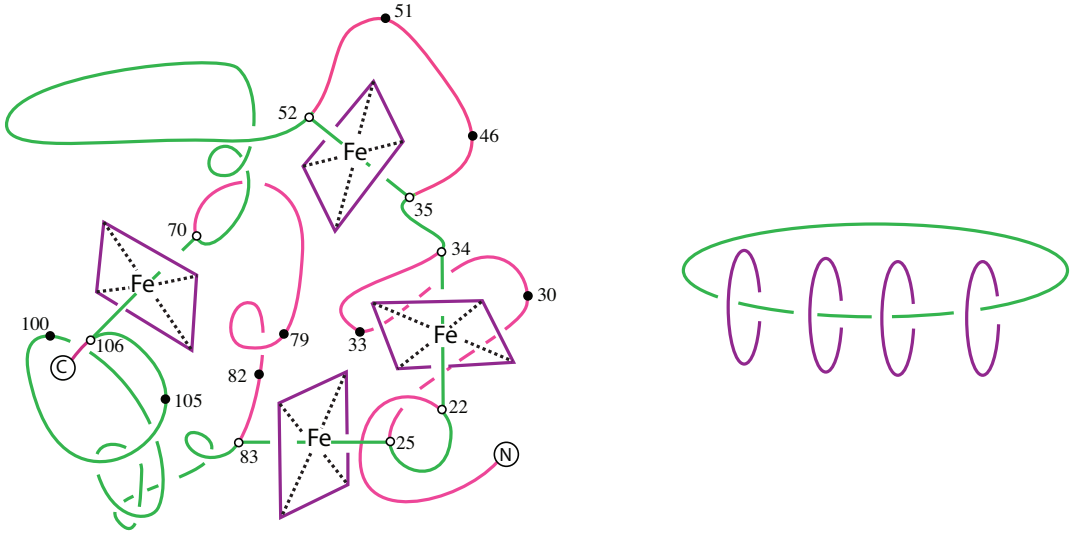

Figure 10: Cytochrome c_3 from *Desulfovibrio vulgarism* Miyazaki and its underlying five-component key ring link.

Perhaps the most fascinating example of linking in protein structures is the linking that exists to protect the bateriophage, HK97. On the exterior of the bacteriophage, 72 rings link together to form a chainmail protective capsid, illustrated in Figure 11. This structure is made up of 12 pentagonal rings and 60 hexagonal

rings linked together to form a spherical shape with icosahedral symmetry. The protective surface is relatively thin, and it is believed that the topological linking adds stability to the capsid. [41, 42].

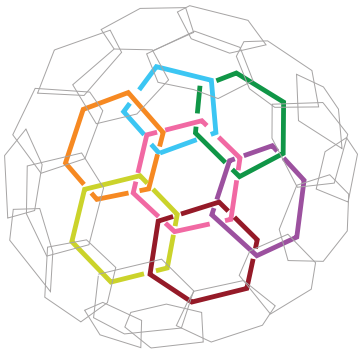

Figure 11: A protein link made up of 12 pentagonal rings and 60 hexagonal rings in a spherical shape with icosahedral symmetry.

Whether a protein link contains two, five, or 72 components, linking seems to greatly increase stability of the resulting structure, likely by reducing the entropy of the unfolded state [42]. As more examples of topological links are discovered in the underlying structures of proteins, it will be fascinating to study the mechanism of their formation.

Spatial Graphs

As we saw above, protein structures can be non-planar because they contain a knot or link. However, these are not the only types of topological non-planarity that can exist within proteins. In particular, non-planar graphs have been found in certain proteins that contain metal clusters. In order to determine whether such a structure contains a non-planar graph, we represent the structure as a graph in three dimensional space where vertices represent atoms or groups of atoms and edges represent bonds. For example, Figure 12 illustrates a spatial graph representing the Fe_4S_4 cluster.

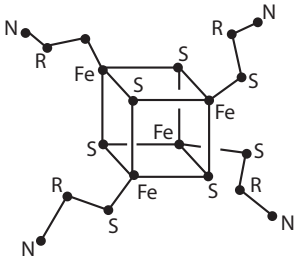

Figure 12: A spatial graph representing the Fe_4S_4 cluster.

For the purpose of identifying metal clusters that prevent a protein structure from being deformed into a plane, we introduce the following two important families of graphs.

Definition 1. *A* complete graph *on n vertices, denoted by K_n, is a graph in which every pair of vertices is connected by an edge.*

We are particularly interested in the complete graph on five vertices K_5 (illustrated in Figure 13) because it is the smallest complete graph which cannot lie entirely in a plane.

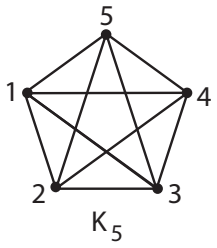

Figure 13: The complete graph on five vertices K_5.

Definition 2. *A* complete bipartite graph $K_{m,n}$ *with two sets of vertices, one containing m vertices and the other containing n vertices, is a graph in which every vertex in the set of m vertices is connected to every vertex in the set of n vertices, but there are no edges connecting a pair of vertices which are both in the same set of vertices.*

We are particularly interested in the complete bipartite graph $K_{3,3}$ (illustrated in Figure 14) because it is the smallest complete bipartite graph of the form $K_{n,n}$ which cannot lie entirely in a plane.

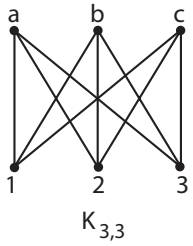

Figure 14: The complete bipartite graph $K_{3,3}$

Observe that the illustrations in Figures 13 and 14 appear as though some edges intersect, which is not the case. We have drawn the graphs in this way because we are focusing on the connectivity between vertices, rather than on a particular conformation of the graph in space. When we talk about a graph exclusively in terms of its connectivity, we refer to it as an *abstract graph*, whereas when we talk about a conformation of a graph in space we refer to the graph as a *spatial graph*. Thus Figures 13 and 14 represent the abstract graphs K_5 and $K_{3,3}$.

If a non-planar graph is contained in a protein structure, it will prevent the protein from being deformed into a plane. However, a non-planar graph can play this role not only if it is a subset of the underlying structure of the protein, but also if it can be obtained from the underlying structure by collapsing edges and/or omitting edges or vertices.

Definition 3. *If a non-planar graph is obtained from a structure by collapsing edges and/or omitting edges or vertices we say that it is a minor of the structure.*

For example, in Figure 15, the graph on the right is obtained from the graph on the left by collapsing the red edges and deleting the blue vertices and blue edge. Thus the graph on the right is a minor of the graph on the left.

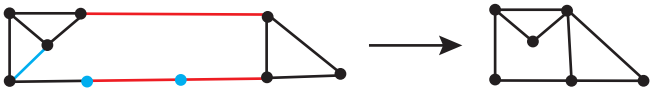

Figure 15: The graph on the right is obtained from the graph on the left by collapsing the red edges and deleting the blue vertices and blue edge.

Mislow and Liang used the following theorem to show that certain protein structures containing cofactors are topologically non-planar [10].

Kuratowski's Theorem. *A graph can lie in a plane if and only if it does not contain either of the graphs K_5 or $K_{3,3}$ as a minor.*

One of Mislow and Liang's examples is *Chromatium* high potential iron protein, which contains the Fe_4S_4 cluster [10]. By Kuratowski's Theorem, showing that the structure contains either a K_5 or $K_{3,3}$ graph as a minor is sufficient to show that the structure is non-planar. The incorporation of the cluster into the backbone ensures, in fact, that it contains both of these graphs as we explain below.

The left image in Figure 16 illustrates the underlying graph of Chromatium high potential iron protein. In order to show that the graph contains $K_{3,3}$ as a minor, we need to identify two sets of three vertices such that every vertex in one set is connected by a disjoint path to every vertex in the other set. One of these sets of vertices (shown as green dots in the middle picture) consists of three of the four iron atoms that occur on the corners of the cube-like metal structure. The other set of vertices (shown as pink dots) consists of two sulfur atoms on the corners of the cube-like structure together with one sulfur atom from a cysteine residue on the polypetide backbone. The orange, black, and purple line segments form disjoint paths from every pink vertex to every green vertex. The light blue vertices are deleted from the middle graph to obtain the $K_{3,3}$ graph. We illustrate $K_{3,3}$ as an abstract graph on the right with the same color scheme to demonstrate that the sets of blue and pink vertices together with the orange, black, and purple edges do indeed form a $K_{3,3}$. It now follows from Kuratowski's theorem that this protein structure is non-planar.

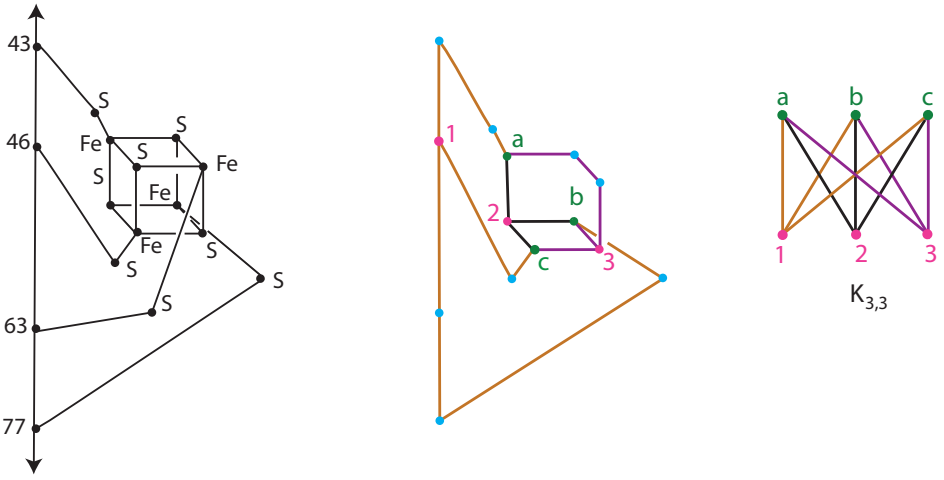

Figure 16: The left graph illustrates the cluster connected to the polypeptide backbone. In the middle graph, six vertices have been partitioned into a green group and a pink group such that there is a path from every green vertex to every pink vertex. Thus the structure contains a $K_{3,3}$ graph, and hence is non-planar.

In Figure 17, we illustrate that the underlying graph of *Chromatium* high potential iron protein also contains K_5 as a minor as a result of the same Fe_4S_4 cluster. In particular, starting with the graph on the left, we collapse the red edges and delete the blue vertices and edges to obtain the graph on the right which has five vertices numbered 1 through 5 and disjoint paths between every pair of vertices. Thus, this cluster also contains K_5 as a minor. This fact alone is also sufficient to show *Chromatium* high potential iron protein is non-planar.

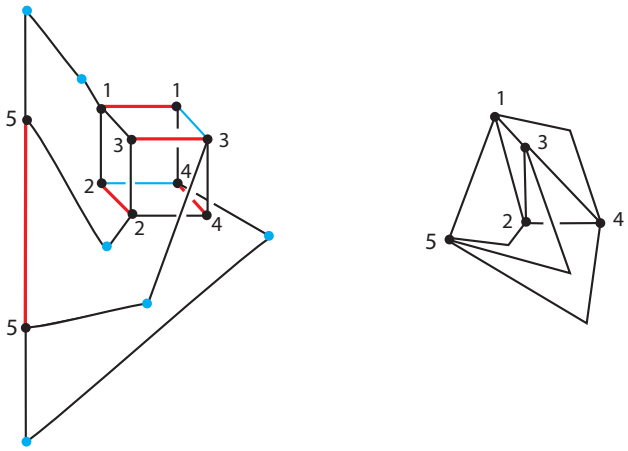

Figure 17: We collapse the red edges and delete the blue vertices and edges to obtain the K_5 graph on the right.

3 Our Results

Non-planarity in Nitrogenase

Nitrogenase is a protein complex that plays a crucial role in the global nitrogen circulation, catalyzing the reduction of atmospheric dinitrogen (N_2) into biologically available ammonia (NH_3). This complex consists of two proteins. The first, often referred to as the Fe protein is a γ_2 dimer with a Fe_4S_4 cluster between the two subunits. The second protein, often referred to as the MoFe protein, is an $\alpha_2\beta_2$-tetramer that contains a total of four complex metal clusters: two P-clusters [Fe_8S_7] (shown on the right in Figure 18) which occur at each α / β subunit interface and two M-clusters [$MoFe_7S_9C$-homocitrate] (shown on the left in Figure 18) which occur within each α subunit. Understanding the structure of these P- and M- clusters had challenged biochemists for decades. In fact, the true elemental composition and connectivity was only established in 2011 [31]. We will show below that as a result of these cofactors, nitrogenase contains several topologically non-planar structures.

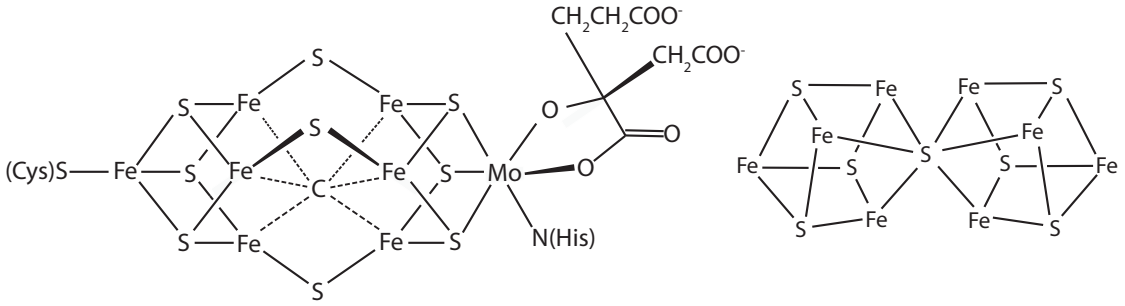

Figure 18: The M- and P- clusters found in nitrogenase shown on the left and right, respectively.

The first metal cluster that we consider is the M-cluster. In order to use Kuratowski's Theorem to prove the non-planarity of nitrogenase, we show that the M-cluster contains a $K_{3,3}$ graph. In particular, in Figure 19, we define one set of three vertices by labeling the top left sulfur atom, the central carbon, and the top right sulfur atom by the letters a, b and c, respectively. We define the second set of three vertices by labeling the top left iron, bottom left iron, and central right iron atoms vertices by the numbers 1, 2, and 3, respectively. We use orange, purple, and green line segments to highlight disjoint paths from vertices a, b, and c to every

numerical vertex. Thus the M-cluster contains a $K_{3,3}$ graph and hence by Kuratowski's theorem, is indeed non-planar.

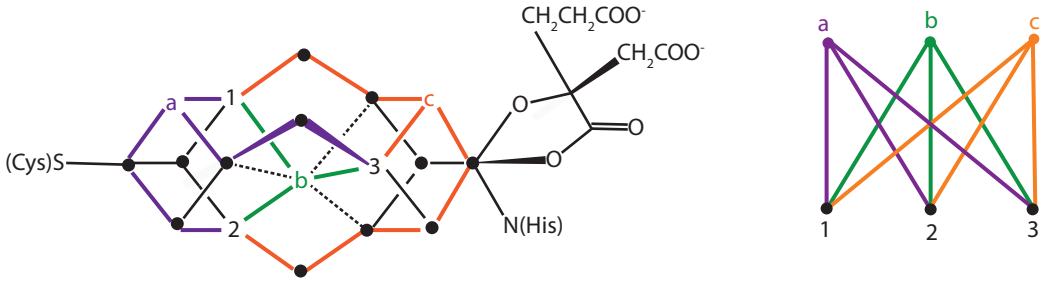

Figure 19: The M-cluster of nitrogenase contains a $K_{3,3}$ graph, and hence is non-planar.

We see as follows that the P-cluster together with the two subunit backbones of nitrogenase contains both of the non-planar graphs $K_{3,3}$ and K_5 as minors. Figure 20 illustrates two different $K_{3,3}$ graphs contained in this structure. Notice that these graphs each make use of part of the pink backbone connected to the cluster. The graph on the left uses the backbone as part of the purple path between vertices a and 2, while the graph on the right uses the backbone as part of the orange path between vertices c and 2. In fact, the P-cluster together with the two backbones contains additional $K_{3,3}$ graphs (not illustrated here).

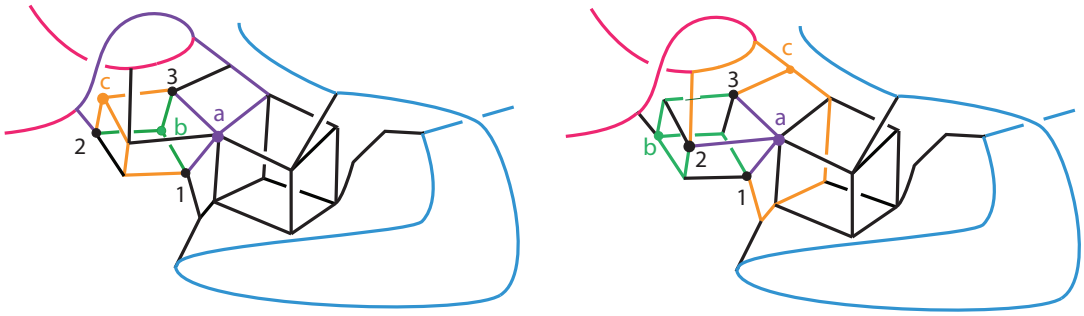

Figure 20: The P-cluster together with the pink backbone contains multiple $K_{3,3}$ graphs.

In order to see that the graph K_5 is also a minor of nitrogenase, we begin with the image on the left in Figure 21. Then we collapse the orange edge and delete the blue vertices, blue edges, and the pink and green arcs of the backbones. Thus we obtain the K_5 illustrated on the right as a minor of the graph on the left. Observe that this K_5 makes use of an arc on the pink backbone in order to join vertex 4 to vertex 5 as well as an arc on the green backbone to join vertex 2 to vertex 5.

Möbius Ladders

Just as a circular path may contain different knots depending on its conformation in space, an abstract graph can have topologically distinct spatial conformations. The conformation of $K_{3,3}$ illustrated on the left side of Figure 22 is known as a *Möbius ladder with three rungs* because of its resemblance to a Möbius strip. While the figure on the left looks quite different from the one on the right, it is easy to check that as abstract graphs the two figures have exactly the same connectivity.

In addition to their intrinsic beauty, Möbius ladders are noteworthy because of the historical role they played in the development of the interdisciplinary field of topological stereochemistry. In particular, in 1961-1962, Ed Wasserman wrote two seminal papers explaining the significance of topology in the analysis of

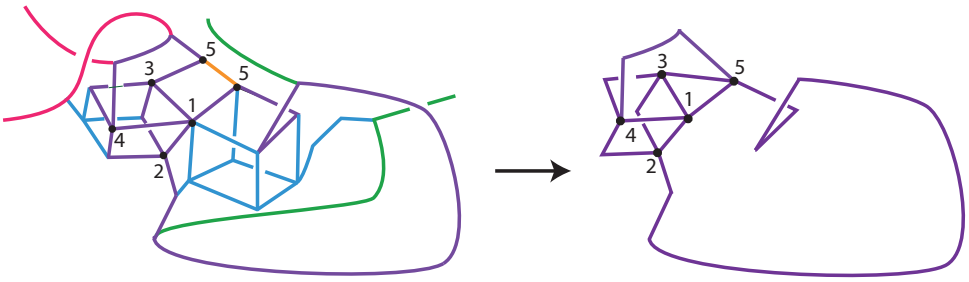

Figure 21: We collapse the orange edge and delete the blue vertices, blue edges, and the green and pink arcs of the backbones to obtain K_5 as a minor of the structure on the left.

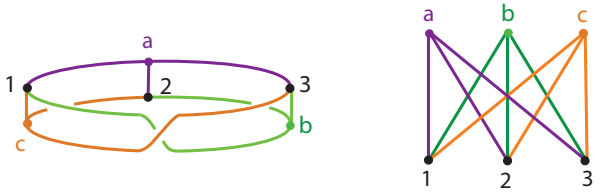

Figure 22: A conformation of $K_{3,3}$ which has the form of a Möbius strip is known as a *Möbius ladder with three rungs*.

molecular structures [5, 39]. In these papers, Wasserman proposed knots, links, and Möbius ladders as future targets for organic synthesis. Total synthesis of the first molecular Möbius ladder was finally achieved by David Walba twenty years later (see Figure 23) [38].

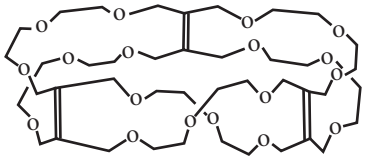

Figure 23: The molecular Möbius ladder synthesized by Walba.

In looking at the NMR spectroscopy data for the Möbius ladder molecule, Walba [37] found that he had to consider non-rigid symmetries as well as the elements of the point group to explain the data he obtained. However, if flexible symmetries are considered, it was not clear whether or not the molecular Möbius ladder could be deformed to its mirror form. In fact, if the Möbius ladder had only two rungs, we could deform it so that it would lie in the plane of the paper as illustrated in Figure 24. The topologist Jon Simon [29] used techniques from knot theory to prove that the molecular Möbius ladder with three rungs was in fact topologically chiral. This marked the beginning of the ongoing interdisciplinary field of topological stereochemistry, which currently includes the study of the topology of DNA, proteins, and other macromolecules.

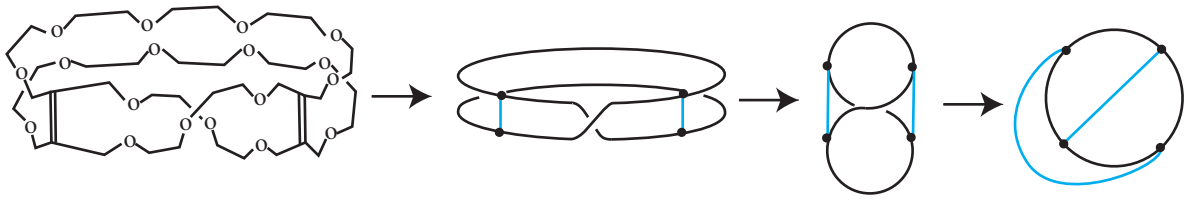

Figure 24: A molecular Möbius ladder with two rungs is topologically planar.

We saw in Figure 19 that the M-cluster of nitrogenase contains the abstract graph $K_{3,3}$. We now see in Figure 25 that the spatial conformation of $K_{3,3}$ in the M-cluster from Figure 19 can actually be deformed to a Möbius ladder with three rungs. We explain the steps of this deformation as follows. Step 1 pulls vertex c to the right and vertex a to the left while deleting the black edges. Step 2 pulls vertex 3 down and towards the right in front of the edge connecting vertices 2 and c. Step 3 rounds out the cycle that goes from vertex 1 to vertex b to vertex 3 to vertex c to vertex 2 to vertex a and then back to vertex 1 so that it looks like a sideways eight. In Step 4, the edge connecting vertices a and 3 is pulled down below the rest of the figure. Step 5 rotates the entire graph by 90° so that the eight is now vertical. In Step 6, vertex 2 is slid under the crossing and up to the left. In the last step the lobes of the eight have been folded together in back of the page to create the usual image of a Möbius ladder.

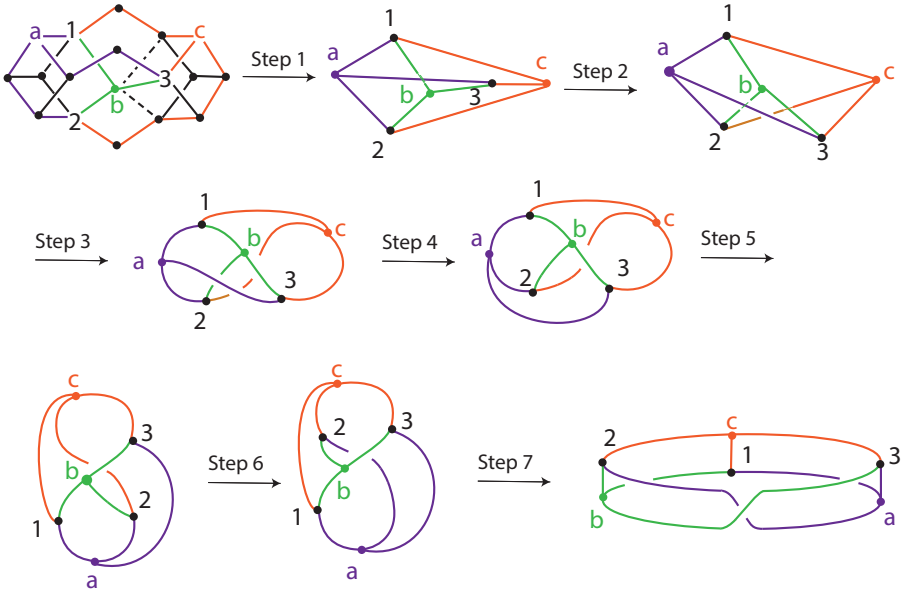

Figure 25: A deformation of the $K_{3,3}$ in the M-cluster of nitrogenase to a Möbius ladder with three rungs.

In fact, Möbius ladders can have any number of rungs, and as long as the number of rungs is at least three the Möbius ladder will necessarily be non-planar. Figure 26 depicts an iron-sulfur cluster connected to the backbone of a protein structure which we show as follows contains a Möbius ladder with four rungs. We begin by illustrating the underlying topological form of the structure on the right in Figure 26.

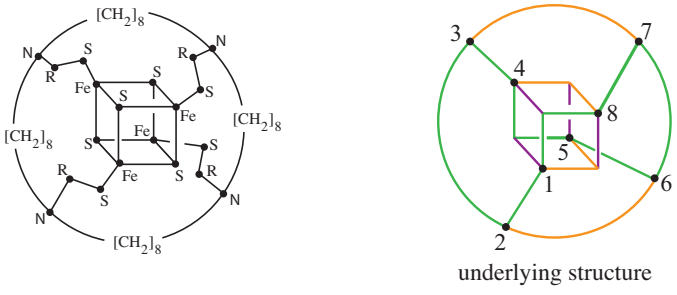

Figure 26: The underlying topological structure of a protein structure with a Fe_4S_4 cluster.

In Figure 27 we omit the four purple edges from the cube-like form in Figure 26. Then, we deform the remaining structure to a Möbius ladder with four rungs.

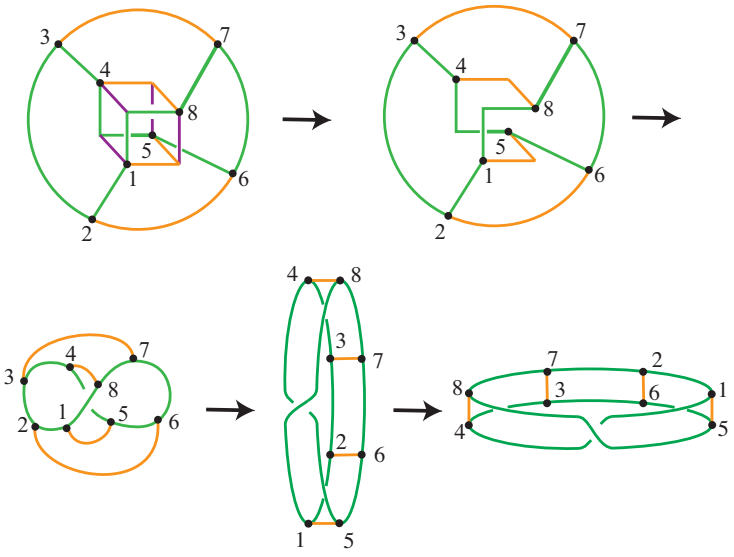

Figure 27: A deformation taking a protein structure with an iron-sulfur cluster to a Möbius ladder with four rungs.

Möbius ladders can also be found in protein structures that do not contain metal clusters. For example, one class of polypeptides that contains Möbius ladders is the family of cyclotides. These small cyclized peptides (typically about 30 amino acids in length) contain many disulfide bonds, and because of their head-to-tail cyclization, we can represent them with a circular backbone [27]. Figure 28 illustrates the highly conserved disulfide connectivity of a cyclotide along with its deformation into a Möbius ladder with three rungs.

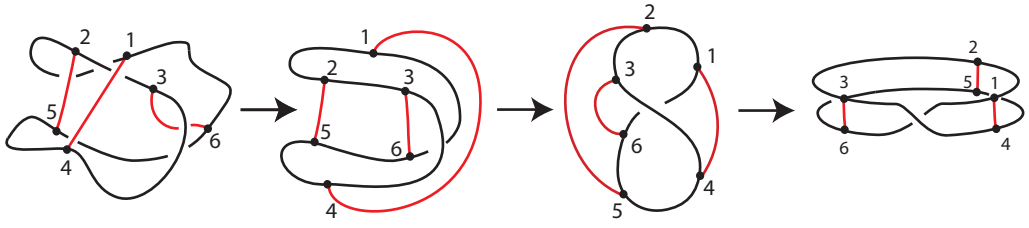

Figure 28: This deformation illustrates that the cyclotide "knot motif" is a Möbius ladder with three rungs.

The image on the left of Figure 28 depicting a typical cyclotide with three disulfide bonds is often described as a "cysteine knot motif." This description seems somewhat inaccurate, since the cyclotide does not actually contain a knot. However, since the cyclotide can be deformed to a Möbius ladder with three rungs, we can interpret a "cysteine knot motif" to mean that the cyclotide has a topologically non-planar underlying structure. Because this connectivity is highly preserved among cyclotides, we can generalize this non-planarity to nearly all cyclotides that contain this connectivity.

Note that cyclotides have been grouped into two structural categories: Möbius cyclotides and bracelet cyclotides. Much less common than the bracelet structure, Möbius cyclotides contain a cis-proline induced backbone twist, making their backbones resemble the edge of a Möbius strip (and hence their name makes sense). However, in fact both the bracelet cyclotides and the Möbius cyclotides contain Möbius ladders with three rungs as illustrated in Figure 28.

Even for proteins that are not cyclized we can identify knots, links and non-planar spatial graphs by using a variation on the topological models to trap knots in open linear protein structures that were introduced by Millett and others [12, 16, 18, 20, 21] (see Section 1). In particular, recall that one approach to modeling a knotted linear protein structure is to place it in the center of a very large ball, and then extend the ends of the

chains out away from the feature of interest and connect them by an arc within the boundary of the ball. We can use a variation on this method to identify the structure depicted on the left in Figure 29 as linked. The green protein has been cyclized as a result of a disulfide bond, shown as a red line segment. By joining the termini of the purple linear protein, we obtain a diagram of a link.

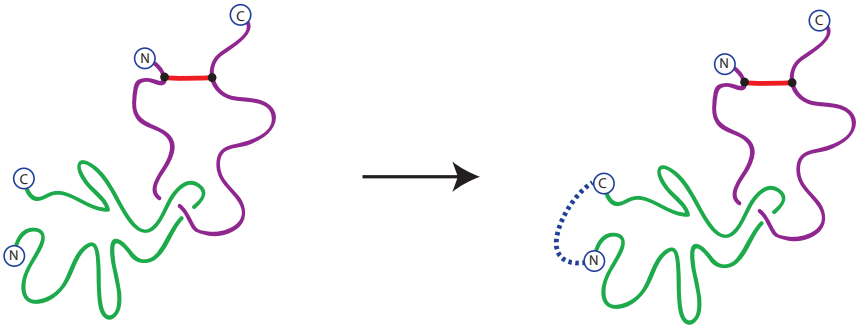

Figure 29: Connecting the termini of the green protein allows us to visualize this structure as a link.

The idea of connecting the N- and C- termini by an arc that is far from any features of interest can also enable us to use Möbius ladders to prove non-planarity of the underlying structures of certain proteins. For example, in Figure 30, we show that the nerve growth factor together with three disulfide bonds can be deformed into a Möbius ladder with three rungs. The first step of the deformation is to deform the edge between vertices 3 and 6 by lifting it over the top arc of the backbone so that it moves from being in front of the backbone to being behind the backbone. The second step is to shorten the edges between vertices 1 and 4 and between vertices 3 and 6 causing the backbone to cross over itself. The third step is to reshape the twisted circle. Finally, in the fourth step we move vertex 2 up and to the right as we shape the twisted circle into the standard form of a Möbius ladder. Like the cyclotide, this protein's disulfide connectivity is also referred to as a "cysteine knot motif", which again can be interpreted as referring to its non-planarity.

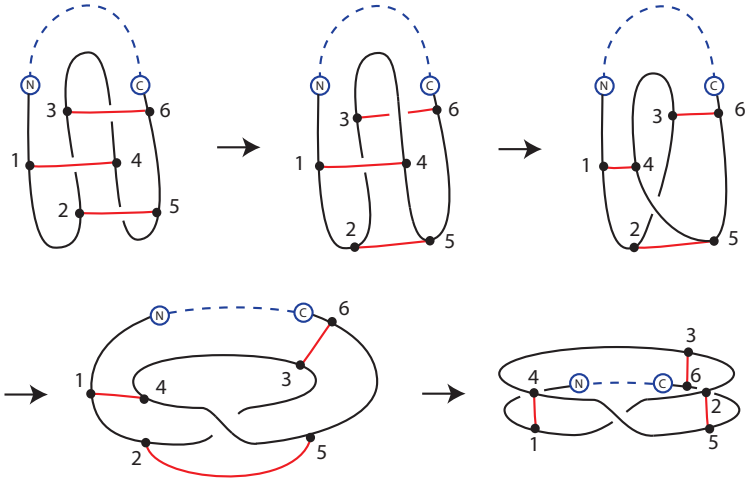

Figure 30: A deformation of the underlying topological structure of the nerve growth factor to a Möbius ladder with three rungs.

The middle steps of the deformations illustrated in Figures 25, 27, 28, and 30 each include a conformation consisting of a circle with three or four edges joining one part of the circle to another. It might be tempting to assume that any such conformation can be deformed into a Möbius ladder. But this is not always the case. To illustrate this, consider the conformation of chymotrypsin inhibitor from Taiwan Cobra, shown in Figure 31.

Observe that in this case we can deform the structure so that it lies in the plane of the paper. In particular, this means that in contrast with Figures 25, 27, 28, and 30, the structure in Figure 31 does not contain a Möbius ladder with three or more rungs.

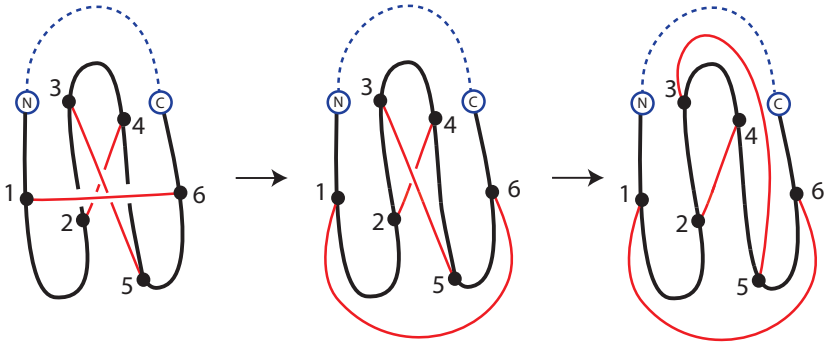

Figure 31: A deformation illustrating that not all proteins with three disulfide bonds contain Möbius ladders. This one, as shown, can be deformed into a plane.

Hypothetical structures

In this section, we propose hypothetical structures illustrating how certain types of non-planar protein structures might occur. First we consider the 5_1 knot, which has not yet been identified in a protein structure. Starting with a protein whose backbone is in the form of a trefoil knot, Figure 32 illustrates how a threading of one end of the backbone through the loop of the trefoil could result in a 5_1 knot. Note that the existence of many protein trefoil knots makes this realization of the 5_1 knot seem possible despite Taylor's argument that the 5_1 knot is an unlikely motif [34].

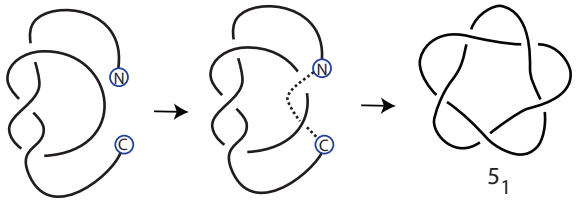

Figure 32: Starting with the trefoil knot, only one threading is required to obtain the 5_1 knot.

Another approach to obtaining a 5_1 knot in a protein structure is to make use of disulfide bonds. In particular, consider a hypothetical protein structure whose polypeptide backbone has five disulfide bonds with connectivity 1-8, 2-5, 3-10, 4-7 and crossings as illustrated in the image on the left in Figure 33. If we ignore every other segment on the backbone, we obtain the second figure, which can then be deformed as illustrated to get the 5_1 knot.

Along these same lines, in Figure 34 we show that a key ring link with three components can hypothetically be found in a single protein which has three disulfide bonds. In particular, consider a hypothetical protein with six cysteine residues on the backbone with disulfide connectivity 1-6, 2-3, 4-5 and crossings as illustrated. In this case (as opposed to that of the 5_1 knot), we need to join the ends of the backbone into a circle as we did in Figure 29 in order to close the blue ring. Then by omitting the black edges of the circle we obtain the three component key ring link as illustrated on the right. Note that this key ring link is different from the one illustrated in Figure 10 because it uses disulfide bonds rather than cofactors.

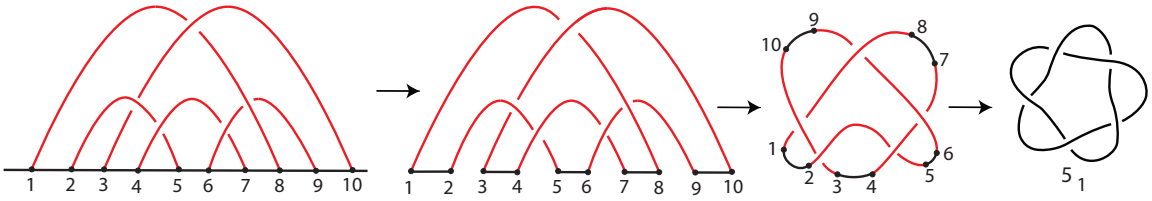

Figure 33: A connectivity that could contain the 5_1 knot.

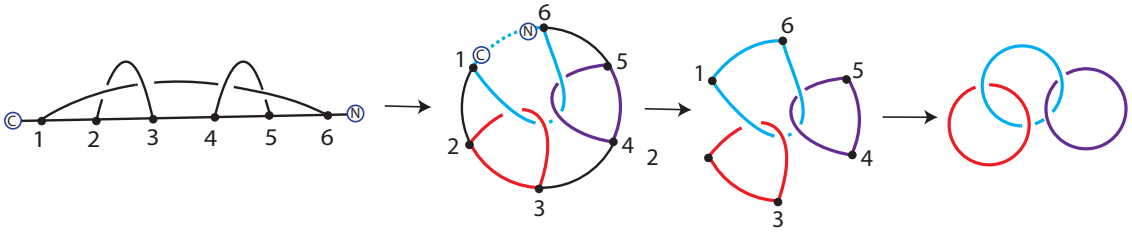

Figure 34: Scheme showing a possible connectivity that would contain a three component key ring link

A similar construction can be used to create a hypothetical protein structure containing the Whitehead link, a motif that has yet to be identified in any protein structure. This link consists of a circular loop that passes through both lobes of a twisted circle, as shown on the right of Figure 35. Such a structure would be particularly noteworthy because the two rings are linked yet have *linking number* equal to zero (the linking number is a standard method of measuring linking used by knot theorists). On the left, we show how this two-component link could hypothetically be obtained from a protein backbone with three disulfide bonds having 1-6, 2-4, 3-5 disulfide connectivity and crossings as shown. As with the structure in Figure 34, we need to join the ends of the backbone into a circle in order to close one of the rings. Again by omitting the black edges of the circle we obtain the desired link.

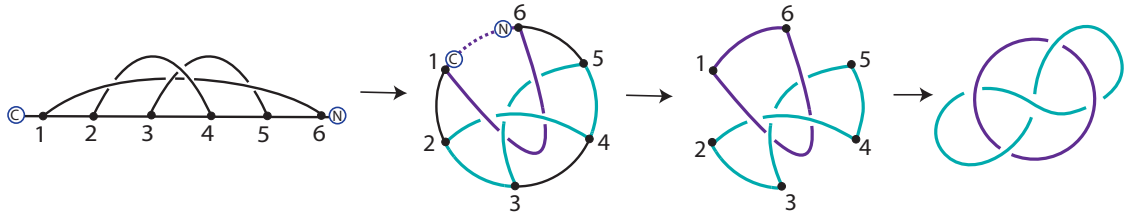

Figure 35: Scheme showing a possible disulfide connectivity that would contain the Whitehead Link.

4 Conclusion

Topologically complex proteins can offer valuable perspectives for understanding protein folding mechanisms and degradation pathways. We have shown that such complexity can be found in protein structures in various ways as the result of knotting, linking, or non-planar graphs. Not only can the polypetide backbone of a protein structure be knotted or a pair of backbones be linked, but knots and links can be found as the result of cofactors and disulfide bonds.

Only four topologically distinct knots have been identified in proteins thus far, and the links that have been identified in proteins are all relatively simple. However, as we saw in our hypothetical structures in

Section 3, there might be many more complex knots and links identified in proteins if the connectivity of disulfide bonds is considered along with the protein backbones.

Finally, we have seen that proteins containing metal structures can include non-planar motifs other than knots or links. In particular, nitrogenase contains several non-planar $K_{3,3}$ and K_5 graphs. Furthermore, Möbius ladders (i.e., ladders in the form of a Möbius strip) can be found in proteins with metal structures including the M-cluster of nitrogenase, as well as in proteins without metal structures including cyclotides. This raises the question of what other interesting spatial graphs might be found in protein structures if cofactors and cysteine disulfide bonds are included in a topological analysis.

Acknowledgement: The first author would like to thank the Institute for Mathematics and its Applications for its hospitality when she was a long term visitor in the autumn of 2013. The authors would also like to thank Dwayne Chambers for technical assistance, as well as Matthew Sazinsky and Helen Wong for helpful discussions.

Conflict of interest: Authors state no conflict of interest.

References

[1] Andrew Belmonte, Michael J Shelley, Shaden T Eldakar, and Chris H Wiggins. Dynamic patterns and self-knotting of a driven hanging chain. *Physical review letters*, 87(11):114301, 2001.

[2] Daniel Bölinger, Joanna I Sułkowska, Hsiao-Ping Hsu, Leonid A Mirny, Mehran Kardar, José N Onuchic, and Peter Virnau. A stevedore's protein knot. *PLoS computational biology*, 6(4):e1000731, 2010.

[3] Zhenbo Cao, Aleksander W Roszak, Louise J Gourlay, J Gordon Lindsay, and Neil W Isaacs. Bovine mitochondrial peroxiredoxin iii forms a two-ring catenane. *Structure*, 13(11):1661–1664, 2005.

[4] Frank B Dean, Andrzej Stasiak, Theo Koller, and Nicolas R Cozzarelli. Duplex dna knots produced by escherichia coli topoisomerase i. structure and requirements for formation. *Journal of Biological Chemistry*, 260(8):4975–4983, 1985.

[5] Harry L Frisch and Edel Wasserman. Chemical topology. *Journal of the American Chemical Society*, 83(18):3789–3795, 1961.

[6] Jack D Griffith and Howard A Nash. Genetic rearrangement of dna induces knots with a unique topology: implications for the mechanism of synapsis and crossing-over. *Proceedings of the National Academy of Sciences*, 82(10):3124–3128, 1985.

[7] Vsevolod Katritch, Wilma K Olson, Alexander Vologodskii, Jacques Dubochet, and Andrzej Stasiak. Tightness of random knotting. *Physical Review E*, 61(5):5545, 2000.

[8] Firas Khatib, Matthew T Weirauch, and Carol A Rohl. Rapid knot detection and application to protein structure prediction. *Bioinformatics*, 22(14):e252–e259, 2006.

[9] Chengzhi Liang and Kurt Mislow. Knots in proteins. *Journal of the American Chemical Society*, 116(24):11189–11190, 1994.

[10] Chengzhi Liang and Kurt Mislow. Topological chirality of proteins. *Journal of the American Chemical Society*, 116(8):3588–3592, 1994.

[11] Chengzhi Liang and Kurt Mislow. Topological features of protein structures: knots and links. *Journal of the American Chemical Society*, 117(15):4201–4213, 1995.

[12] Rhonald C Lua and Alexander Y Grosberg. Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS computational biology*, 2(5):e45, 2006.

[13] Anna L Mallam and Sophie E Jackson. A comparison of the folding of two knotted proteins: Ybea and yibk. *Journal of molecular biology*, 366(2):650–665, 2007.

[14] Marc L Mansfield. Are there knots in proteins? *Nature Structural & Molecular Biology*, 1(4):213–214, 1994.

[15] Gurvan Michel, Véronique Sauvé, Robert Larocque, Yunge Li, Allan Matte, and Miroslaw Cygler. The structure of the rlmb 23s rrna methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure*, 10(10):1303–1315, 2002.

[16] Kenneth C Millet and Benjamin M Sheldon. Tying down open knots: A statistical method for identifying open knots with applications to proteins. In Jorge A Calvo, Kenneth C Millet, Eric J Rawdon, and Andrzej Stasiak, editors, *Physical and numerical models in knot theory. Singapore: World Scientific*, pages 203–217. 2005.

[17] Kenneth Millett, Akos Dobay, and Andrzej Stasiak. Linear random knots and their scaling behavior. *Macromolecules*, 38(2):601–606, 2005.

[18] Kenneth C Millett. Physical knot theory: an introduction to the study of the influence of knotting on the spatial characteristics of polymers. In *Introductory Lectures on Knot Theory: Selected Lectures Presented at the Advanced School and Conference on Knot Theory and Its Applications to Physics and Biology*, volume 46, pages 346–378. World Scientific, 2011.

[19] Kenneth C Millett, Eric J Rawdon, Andrzej Stasiak, Joanna I Sułkowska, et al. Identifying knots in proteins. *Biochemical Society Transactions*, 41(part 2):533–537, 2013.

[20] Eleni Panagiotou, Kenneth C Millett, and Sofia Lambropoulou. Quantifying entanglement for collections of chains in models with periodic boundary conditions. *Procedia IUTAM*, 7:251–260, 2013.
[21] Eleni Panagiotou, Christos Tzoumanekas, Sofia Lambropoulou, Kenneth C Millett, and Doros N Theodorou. A study of the entanglement in systems with periodic boundary conditions. *Progress of Theoretical Physics Supplement*, 191:172–181, 2011.
[22] Piotr Pierański, Sylwester Przybył, and Andrzej Stasiak. Tight open knots. *The European Physical Journal E*, 6(2):123–128, 2001.
[23] Elizabeth Pleshe, John Truesdell, and Robert T. Batey. Structure of a class II TrmH tRNA-modifying enzyme from *Aquifex aeolicus*. *Acta Crystallographica Section F*, 61(8):722–728, Aug 2005.
[24] Raffaello Potestio, Cristian Micheletti, and Henri Orland. Knotted vs. unknotted proteins: evidence of knot-promoting loops. *PLoS computational biology*, 6(7):e1000864, 2010.
[25] Eric J Rawdon, Kenneth C Millett, Joanna I Sułkowska, Andrzej Stasiak, et al. Knot localization in proteins. *Biochemical Society Transactions*, 41(part 2):538–541, 2013.
[26] Dale Rolfsen. *Knots and links*. Publish or Perish Inc., Berkeley, Calif., 1976. Mathematics Lecture Series, No. 7.
[27] K Johan Rosengren, Norelle L Daly, Manuel R Plan, Clement Waine, and David J Craik. Twists, knots, and rings in proteins structural definition of the cyclotide framework. *Journal of Biological Chemistry*, 278(10):8606–8616, 2003.
[28] Chris Sander and Reinhard Schneider. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins: Structure, Function, and Bioinformatics*, 9(1):56–68, 1991.
[29] Jonathan Simon. Topological chirality of certain molecules. *Topology*, 25(2):229–235, 1986.
[30] Jonathan Simon. Long tangled filaments. *Applications of Knot Theory: American Mathematical Society, Short Course, January 4-5, 2008, San Diego, California*, 66:155, 2009.
[31] Thomas Spatzal, Müge Aksoyoglu, Limei Zhang, Susana LA Andrade, Erik Schleicher, Stefan Weber, Douglas C Rees, and Oliver Einsle. Evidence for interstitial carbon in nitrogenase femo cofactor. *Science*, 334(6058):940–940, 2011.
[32] Fusao Takusagawa and Shigehiro Kamitori. A real knot in protein. *Journal of the American Chemical Society*, 118(37):8945–8946, 1996.
[33] William R Taylor. A deeply knotted protein structure and how it might fold. *Nature*, 406(6798):916–919, 2000.
[34] William R Taylor. Protein knots and fold complexity: some new twists. *Computational biology and chemistry*, 31(3):151–162, 2007.
[35] EJ Janse Van Rensburg, DAW Sumners, E Wasserman, and SG Whittington. Entanglement complexity of self-avoiding walks. *Journal of Physics A: Mathematical and General*, 25(24):6557, 1992.
[36] Peter Virnau, Leonid A Mirny, and Mehran Kardar. Intricate knots in proteins: Function and evolution. *PLoS computational biology*, 2(9):e122, 2006.
[37] David M Walba. Topological stereochemistry. *Tetrahedron*, 41(16):3161–3212, 1985.
[38] David M Walba, Rodney M Richards, and R Curtis Haltiwanger. Total synthesis of the first molecular möbius strip. *Journal of the American Chemical Society*, 104(11):3219–3221, 1982.
[39] Edel Wasserman. Chemical topology. *Scientific American*, 207:94–102, 1962.
[40] Steven A Wasserman, Jan M Dungan, and Nicholas R Cozzarelli. Discovery of a predicted dna knot substantiates a model for site-specific recombination. *Science*, 229(4709):171–174, 1985.
[41] William R Wikoff, Lars Liljas, Robert L Duda, Hiro Tsuruta, Roger W Hendrix, and John E Johnson. Topologically linked protein rings in the bacteriophage hk97 capsid. *Science*, 289(5487):2129–2133, 2000.
[42] Todd O Yeates, Todd S Norcross, and Neil P King. Knotted and topologically complex proteins as models for studying folding and stability. *Current opinion in chemical biology*, 11(6):595–603, 2007.